

<https://helda.helsinki.fi>

---

## Higher-Order Defeat and Evincibility

Lasonen-Aarnio, Maria

Oxford University Press  
2019

---

Lasonen-Aarnio , M 2019 , Higher-Order Defeat and Evincibility . in M Skipper & A  
Steglich-Peterson (eds) , Higher-Order Evidence: New Essays . Oxford University Press . <https://doi.org/10.1093/oso>

---

<http://hdl.handle.net/10138/336859>

<https://doi.org/10.1093/oso/9780198829775.003.0007>

---

acceptedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# Higher-Order Defeat and Evincibility

Maria Lasonen-Aarnio

*Final version, to appear in Higher-Order Evidence: New Essays*

*(Skipper & Steglich-Petersen eds., OUP)*

## Abstract

One of the ambitions of the past decades of epistemology has been to accommodate the view that ‘higher-order evidence’ that a belief one holds is rationally flawed has a systematic kind of defeating force with respect to that belief. Such a view is committed to two claims. First, it is possible to acquire misleading evidence about the normative status of one’s doxastic states: even if one’s belief is perfectly rational, one might acquire deeply misleading evidence that it is irrational. Second, such evidence has defeating force with respect to the belief, zapping its rational status. My aim is to do two things. First, I outline a view I call *normative evincibility*, according to which one always has a kind of epistemic access – access that can come in different strengths – to the normative status of one’s doxastic states (intentions, actions, etc.). I show how commitment to higher-order defeat in effect incurs a commitment to a form of normative evincibility. Second, I argue that the idea that it is possible to acquire misleading evidence about the normative status of one’s doxastic states is in tension with evincibility. Hence, there is a deep tension inherent in views committed to a systematic phenomenon of defeat by higher-order evidence.

## I. The plan

Here is a package of views featured notably in debates about higher-order evidence. First, it is possible to acquire higher-order evidence that I have made some sort of rational error in assessing my evidence and hence, that my doxastic states fall short of a normative standard such as rationality. Second, such evidence has defeating force: sufficiently strong evidence that a belief is irrational defeats the rationality of that belief. As a result, the evidence calls for revising my opinions. For instance, if I become reasonably confident that my belief in *p* is irrational, I ought to suspend judgment about whether *p*.<sup>1</sup> Views calling for such revision are often motivated by appeal to an ideal of epistemic modesty.

“...we may make mistakes in assessing evidence...reason to suspect that we’ve made a mistake in assessing the evidence is often also reason to be less confident in the conclusion we initially came to. The rationale for revision, then, expresses a certain kind of epistemic modesty” (Christensen 2013: 77)

In what follows, my first aim will be to establish a claim that I think is almost self-evident once spelled out: that commitment to a systematic kind of defeat entails commitment to an epistemic access condition I call *normative evincibility*. My second aim will be to argue that there is a deep tension inherent in the view that while we can acquire misleading evidence regarding the normative status of our beliefs, their normative status is always evincible – that is, there is a deep tension inherent in views committed to defeat by higher-order evidence. I hope, then, to argue against the above package of views by providing a kind of *reductio*.

Several authors (including myself) have pointed out that those who endorse higher-order defeat, and the kinds of ‘level-connecting principles’ that go with it, appear to be committed to ruling out the possibility of certain kinds of epistemic situations that arguably *can* arise.<sup>2</sup> But the

---

<sup>1</sup> See, for instance, Bergmann (2006), Christensen (2007a, 2010a), Elga (*unpublished*), Feldman (2005) Horowitz (2014), Schechter (2011).

<sup>2</sup> See Christensen (2010b), Williamson (2011), Horowitz (2014), Worsnip (2018), Lasonen-Aarnio (*forthcoming*)

discussions so far leaves open the possibility of treating these seeming counterexamples as isolated, peculiar cases in which exceptions to generalizations are called for. My hope is to shift the dialectic: the kinds of cases that have been discussed are surface symptoms of a deeper, more endemic problem.

I will begin by saying how I think of higher-order evidence, and by spelling out in more detail two distinct commitments of views endorsing higher-order defeat (section 2). I then formulate a kind of epistemic access condition I call *evincibility*, arguing that endorsing higher-order defeat commits one to endorsing the evincibility of rationality, a kind of epistemic access to the rational status of one's (present) doxastic states (section 3). Finally, I discuss the internal coherence of views committed to defeat by higher-order evidence (section 4).

## 2. Higher-Order Evidence: A General Characterization

'Higher-order evidence' (or 'HOE') is mostly epistemologist talk, understood in terms of an array of cases now somewhat canonical in the literature. Here is a case I take to be canonical enough:

### **Resident**

Rezi is a medical resident in charge of diagnosing a patient and prescribing the appropriate treatment. After carefully reflecting on the patient's symptoms, labs, and other relevant information, she becomes confident, and comes to believe, that the appropriate treatment is a 10mg dose of *Wellstrol*. Rezi knows that due to her constantly sleep-deprived state, rarely an isolated cognitive *blip* will occur: an error in her reasoning that results in her arriving at a random conclusion by a cogent-seeming process. Rezi typically cannot detect such blips herself. She also knows that just when such blips occur is itself random. As it happens, Rezi's performance is being monitored by a team of neuroscientists who can see the fine cognitive workings of her mind. As she is about to prescribe *Wellstrol* for her patient, the neuroscientists inform her that a blip occurred: her diagnosis is the output of a process no better than a random guess at tracking the evidentially supported opinion. Though the neuroscientists are all but infallible, this time they are mistaken: Rezi's original reasoning was impeccable, and a 10mg dose of *Wellstrol* was appropriate given her evidence.

The testimonial evidence that a blip occurred is higher-order evidence. Let me, however, flag at the outset that I do not want to imply that any body of evidence can be partitioned into two parts, the first-order evidence and the higher-order. It is better to think about higher-order evidence in terms of higher-order *import*. A single piece of evidence can have import regarding both first-order questions and higher-order questions. (And a piece of evidence can have different kind of import in different situations, depending on other evidence one has.) What, then, is it for a piece of evidence to have higher-order import in a given situation?

According to one popular slogan, higher-order evidence is *evidence about evidence*, or evidence bearing on evidential relations.<sup>3</sup> Taking this view as a starting point, we could say that to have higher-order import is to bear on propositions regarding what one's evidence supports. Applied to the above case, the suggestion would be that Rezi's evidence regarding her cognitive blip now makes it less likely that her original medical evidence supported (made likely) that a 10 mg dose of *Wellstrol* is the appropriate treatment. But at the very least, I think this is far from clear. Evidence that her opinion was the result of a blip is evidence that her opinion is the output of a random process. But the fact that a random process produced the belief that a 10 mg dose of *Wellstrol* is the appropriate treatment has no bearing either on whether *Wellstrol* is the appropriate treatment, *nor* on whether it was likely on Rezi's original evidence that it is appropriate. It does, however, have bearing on whether Rezi's belief that *Wellstrol* is the

---

<sup>3</sup> One of the earliest proponents this idea was Feldman (2005).

appropriate treatment has, or ever had, epistemic statuses such as being *rational*, being *justified*, or constituting *knowledge*.

Thinking about HOE exclusively as evidence about evidential relations blurs a more general phenomenon from view, one that is not specific to the epistemic realm. On a broader characterization that I favour, higher-order import is a matter of bearing on a relevant normative status of one's mental states, or even of one's actions. In the epistemology literature focus has been on evidence that bears on normative statuses such as rationality, justification, and perhaps knowledge. More generally, I suggest we think of HOE as any evidence bearing on whether one's beliefs (intentions, actions) are permitted in some relevant sense, and hence, on whether they have some relevant normative status. There are two ways in which evidence might *bear on* whether, say, a belief has the status of being rational: the evidence might make it more likely that the belief has the status, or it might make it less likely. This talk of making likelier or less likely should be understood against the background of the total evidence one already has: if one thinks that updating happens by conditionalization, then what it is for a piece of evidence  $e$  to bear on a proposition  $p$  is for it to be the case that  $\Pr(p | e) \neq \Pr(p)$  (here I am assuming that 'Pr' denotes something like evidential probabilities).

I will leave it largely open what restrictions (if any) should be placed on the normative statuses that HOE can bear on. I will assume that there is at least one sense of *permissibility* (and correspondingly, of being required and forbidden) pertaining to doxastic states, that there is a range of correct norms that concern permissibility in this sense, that evidence bearing on such permissibility counts as HOE, and that, according to views committed to defeat by HOE, sufficiently strong misleading evidence bearing on the permissibility of at least a range of doxastic states (such as beliefs) defeats their permissibility. Epistemologists typically talk of the normative statuses of *rationality* or *justification* in ways that fit this bill: a doxastic state is permitted in the relevant sense just in case it is rational or justified. Though the discussion below could be conducted using any term picking out a normative status with the desired role, I will use 'rational'. I will use 'irrational' for a status had by a state that is not rational. (This might be a simplification, in so far as one thinks it is possible for a doxastic state to not be rational without being irrational, but I doubt anything of substance rests on it.) Hence, I will use 'rational' in ways that epistemologists commonly do. However, there is also another way of reading most of the rationality-talk below, namely, as a placeholder for any genuinely normative status of a doxastic state, such that evidence that a doxastic state lacks that status counts as HOE. By using the word 'rational', I do not intend to commit myself to any form of internalism, or even to the indispensability of the ideology of rationality.

My characterization of HOE is somewhat schematic. But this broader characterization allows us to see parallels between various questions that have been discussed in different areas of philosophy. Before looking at these parallels, I will outline two commitments of a view on which there is a systematic phenomenon of higher-order defeat: the first is an existence claim I have implicitly assumed, a commitment to the possibility of acquiring even deeply misleading HOE; the second is a commitment to a systematic kind of defeat by such evidence. (I will set aside the further commitment to a positive recommendation regarding how one ought to adjust one's doxastic states when faced with HOE.)

(I) Acquiring evidence is a diachronic process: I start out with an initial body of evidence, and then acquire new evidence calling me to update my doxastic states. For instance, I may start out confident that a given belief is rational, but as a result of acquiring new evidence, it may then become unlikely on my new total evidence that my belief was (or, assuming that I still hold

it, is) rational.<sup>4</sup> What I take to be the central existence question regarding HOE concerns the possibility of acquiring deeply misleading evidence bearing on the normative status of one's beliefs, actions, etc. Applied to the normative status of rationality, we can formulate this existence claim as follows:

#### *Acquisition*

It is possible to acquire evidence making it rational to believe, or at least be reasonably rationally confident, in falsehoods regarding the rational status of one's doxastic states.<sup>5</sup>

Most of the literature on HOE in epistemology assumes *Acquisition*: I might hold a perfectly rational belief, but then acquire misleading evidence – whether in the form of peer disagreement, evidence that I am susceptible to cognitive biases, that I have been given a reason-distorting drug, etc. – that my belief is irrational, evidence that makes it rational for me to now be at least reasonably confident in the relevant falsehood.<sup>6</sup>

It is worth noting that some authors have denied the existence of at least a certain kind of misleading HOE, arguing that a deep kind of rational uncertainty about normative (as opposed to descriptive) matters is impossible. For instance, Elizabeth Harman (2011: 460-462) claims that false moral beliefs that arise not from ignorance of non-moral facts, but from ignorance of moral facts, are not epistemically justified. Mike Titelbaum (2015) argues that what he describes as *a priori* truths about what rationality requires can never be too unlikely on one's overall evidence, for there is always *a priori* evidence in place for those truths.<sup>7</sup> These views only deny that it is possible to acquire HOE that is misleading in virtue of pointing to false *normative* claims. But numerous discussions of defeat assume, for instance, that one can acquire even radically misleading evidence regarding support-facts, facts about what a body of evidence confirms in the absolute sense, or facts about how likely various propositions are on a body of evidence.<sup>8</sup> And the whole discussion of normative uncertainty (which I mention below), is propelled by

---

<sup>4</sup> There are delicate issues of time-indexing that arise here. Distinguish between the condition *that my belief is rational*, which can be true at one time and false at another, and the proposition *my belief is rational at t*. Upon acquiring HOE bearing negatively on the rationality of one's belief, presumably the condition *that my belief in p is rational* becomes less likely. However, in such cases HOE is also supposed to have a kind of retrospective bearing on whether it was ever rational for one to believe *p* in the first place (cf Christensen 2010a). Hence, the proposition *my belief is rational at t* also becomes less likely. Note that this sort of thing often happens when we get evidence bearing on states that extend through time. For instance, I might think that I am angry at someone, but get evidence that I am not in fact angry, just disappointed. This is evidence bearing on the condition *that I am angry*, but it also bears on whether I was ever angry to begin with.

<sup>5</sup> Note that by a 'doxastic state' I mean a state such as believing a proposition *p*, disbelieving *p*, or suspending judgment in *p*. I do not mean an entire state consisting of propositional attitudes to many different propositions.

<sup>6</sup> Christensen (2013: 90) dubs a similar claim *Respecting evidence of our epistemic errors*: "This sort of ideal requires, for example, that in typical cases where one is initially confident that *p*, and one encounters good evidence that one's initial level of confidence in *p* is higher than that supported by one's first-order evidence ... , one will give significant credence to the claim that one's initial level of credence is too high. This sort of requirement applies even when one hasn't actually made an error in one's initial assessment of the evidence."

<sup>7</sup> Littlejohn (2018) holds a similar view, though on different grounds. If, for instance, *R* is a reason to believe *p*, and I have *R*, but I fail to believe *p*, then I manifest *de re* unresponsiveness to my reasons. But if I falsely believe that *R* is not a reason to believe *p*, Littlejohn thinks I manifest the same kind of insensitivity to my reasons. (This is so even if I have arbitrarily strong evidence for the false view – Littlejohn denies that rationality is a matter of evidential support.) Such a view reflects remarks made by Harman (2011: 460): "Believing a certain kind of behaviour is wrong on the basis of a certain consideration is a way of caring about that consideration".

<sup>8</sup> For instance, in Feldman's (2005) example, a student acquires misleading evidence that her perceptual experience that *p* does not provide good support for believing *p*. See also Christensen (2010a) for uncertainty regarding logical entailment, Horowitz (2013) for a case involving uncertainty about what a body of evidence supports.

the assumption that we often find ourselves in situations of even a deep kind of rational uncertainty about what normative (e.g. ethical) theory is correct.<sup>9</sup>

Let me now look at commitment to the defeating force of the relevant kind of HOE.

(2) By views committed to defeat by higher-order evidence I mean views on which sufficiently strong HOE that one falls short of a relevant normative status more or less always, and necessarily, defeats the relevant normative status.

I myself am sympathetic to a view on which normative statuses like knowledge and rationality are *sometimes* lost as one acquires HOE. However, the reasons for this have nothing to do specifically with HOE. For instance, my confidence in a proposition *p* might be rational to begin with, in virtue of the fact that *p* is likely on my evidence. I might then acquire evidence that bears on *both* the rationality of my confidence in *p*, and on *p* itself. If the result of such evidence is that *p* is no longer likely, then it is no longer rational for me to be confident in *p*. Or, my belief in *p* might constitute knowledge to start out with, but the result of acquiring HOE is that I no longer know *p*, because some condition on knowledge such as safety is no longer satisfied. Perhaps, for instance, I re-base my belief in such way that it is no longer safe from error. However, I very much doubt that such effects are systematic: sometimes HOE has no evidential bearing on the relevant first-order propositions, and sometimes one can continue to know despite the HOE (e.g. Lasonen-Aarnio 2010, 2014). In fact, I think it is plausible that the *Resident* case described above is like this, for evidence that a blip occurred simply has no bearing on whether Wellstrol is the appropriate treatment. And I am inclined to think that Rezi *can* retain her knowledge, which is not to say that she is immune to criticism if she retains her belief. But here I won't rely on these ideas: I bring them up to distinguish between views that endorse a (systematic) phenomenon of higher-order defeat, and views that do not.

Further, not any view on which there is always *something* wrong with retaining a doxastic state despite having evidence that it is irrational is committed to what I am referring to as 'higher-order defeat'. For instance, one might think that there is something *blameworthy* about Rezi if she retains her belief, despite having strong evidence that it is the result of a more or less random cognitive process. One might think that there is something *incompetent* or *unvirtuous* about it; or that it fails to reflect a good *strategy* for having beliefs that constitute knowledge, that are proportioned to the evidence, or that are rational. Closely related is the thought that it manifests some *bad dispositions*, dispositions that tend to lead one astray. Indeed, I think that Rezi manifests some bad dispositions, for given natural assumptions about her psychology, she cannot be disposed to retain her belief *only when* the higher-order evidence is misleading. But such an answer does not commit one to the idea that the relevant epistemic statuses are in fact defeated – evidence that her belief is irrational need not zap its status as rational, or its status as knowledge.

For a normative status **N**, we can ask whether HOE bearing in a specific way on whether one's mental state or action has **N** has a systematic kind of defeating force with respect to that status. For instance, does evidence bearing in this way on whether an action is morally right defeat the moral rightness of the action? Does evidence bearing in this way on whether a belief is rational defeat its rationality? As we will see, there are disputes about exactly what this bearing should be. I will work with a simple view that explains common verdicts across standard candidate cases of higher-order defeat.

---

<sup>9</sup> See also e.g. Littlejohn (2018) for considerations in support of the idea that one's evidence could provide arbitrarily strong support for a false normative view.

Here is a pretty standard form that such cases take.<sup>10</sup> Initially, one rationally believes a proposition  $p$ , or is rationally confident in  $p$ , based on evaluating some body of evidence, or of performing some reasoning. One then acquires new evidence the result of which is that it becomes reasonably likely that the doxastic process producing one's belief was flawed and hence, that one's belief or high confidence in  $p$  was irrational.<sup>11</sup> Typically, this is now likely to degree .5 (as in cases of peer disagreement) or above. On the view under consideration, then, there is a threshold  $r$  such that even if a state was in fact rational, its becoming likely to degree  $r$  or above that a state is irrational defeats the rationality of the state, rendering it irrational:

#### *Higher-Order Defeat*

Evidence making it likely to (at least) degree  $r$  or above ( $0 < r < 1$ ) that a relevant doxastic state is irrational defeats the rationality of that state.<sup>12</sup>

I will remain largely noncommittal about what the relevant threshold  $r$  is. One suggestion is that  $r$  is any value above .5: evidence that make it *likely* that a state is irrational defeats the rationality of that state. But note that in some examples of higher-order defeat it is not even .5 likely that one has committed a rational error<sup>13</sup>. Further, on such a view it would be difficult to make sense of conciliatory verdicts regarding cases of peer disagreement: in typical cases, a subject thinks it is equally likely that she committed a rational error as that her peer did. And there are numerous other examples in the literature besides in which defeat is assumed to occur in a situation in which the higher-order evidence makes it merely .5 likely that one's initial belief is irrational.<sup>14</sup>

Paradigmatically, higher-order defeat is assumed to apply to belief: a belief cannot be rational if it is too likely on one's evidence that it irrational. One might think that is true, but deny a corresponding thesis for suspension, or for credences (assuming such states are psychologically real): perhaps it can, for instance, be rational to assign a middling credence to  $p$ , even if it is likely that a middling credence is irrational.<sup>15</sup> It won't matter much whether

<sup>10</sup> E.g. Christensen (2007b, 2010a), Elga (2007: 484), Feldman (2005), Kelly (2010), Lackey (2010), White (2009: 237) Horowitz (2014: 719). Not all of these authors defend defeat by HOE; the point is that examples they discuss have this structure.

<sup>11</sup> A somewhat natural interpretation of the evidential dynamics of standard higher-order defeat scenarios is that it *becomes* likely to some sufficiently high degree that a given doxastic state one is in is irrational. This involves confirmation in both a comparative and absolute sense. However, I will assume that the (putative) defeating force of the new evidence derives from the fact that it is now sufficiently likely in the absolute sense that one's belief is irrational, or that it is now rational for one to believe, or to be confident, that the belief is irrational. That it is now *more* likely than before that a given doxastic state is irrational follows from this assumption, together with the assumption that the state was previously rational (note that many descriptions of scenarios involving defeat by HOE explicitly assume this; as an example, see the *Sleepy Detective* case in Horowitz 2014).

<sup>12</sup> Parallel questions can be asked in a more propositional mode. For instance: 'Is (strong) evidence that it is not (propositionally) rational for one to believe  $p$  compatible with its being (propositionally) rational for one to believe  $p$ ?' I will focus on doxastic normative statuses. On some views, defeat of a doxastic status like doxastic justification does not entail defeat of a propositional status such as propositional justification (see, for instance, van Wietmarschen 2013).

<sup>13</sup> For instance, in Sliwa and Horowitz's (2015) example an anaesthesiologist should still think it is 60% likely that his initial judgment was correct.

<sup>14</sup> E.g. examples motivating calibrationism, like Schoenfield's (2015) pilot who learns that she is at risk of hypoxia.

<sup>15</sup> An alternative way of thinking works in terms of *expectations*. Some metanormativists have defended the view that one ought (it is rational to, it is morally right to, etc.) to perform the act with the highest expected value (e.g. Ross 2006); others have argued that one ought to go for the act with the highest expected choiceworthiness (e.g. MacAskill 2014). Similarly, in the epistemology literature some have defended *Rational Reflection* -type principles: one ought to adopt the credence that is one's (rational) expectation of the rational credence, or some refinement of this though (Christensen 2010b, Elga 2013, Dorst *forthcoming*). This paper does not take on such principles (see

*Higher-Order Defeat* is assumed to apply to *all* doxastic states, as long as it applies to some. In what follows, I will assume that it applies *at least* to the state of belief. Note also that on a possible view of defeat, evidence as such that a state is normatively flawed is not sufficient for defeat; what constitutes a defeater is a belief, or sufficiently high degree of confidence, based on such evidence. Whether it is evidence as such, or doxastic states based on such evidence, that act as defeaters, won't matter, for the main dialectic of this paper can be played out on either view.

Though the term 'higher-order evidence' is rarely used outside epistemology, the idea of defeat by higher-order evidence has clear repercussions in discussions of practical reason and ethics. Consider, for instance, Joseph Raz's (1975) discussion of what he calls exclusionary reasons.<sup>16</sup> Raz argues that reasons to think one is incapacitated (drunk, fatigued, emotionally upset, etc.) can exclude certain first-order reasons for action from consideration, thus defeating them. Consider Ann, who has excellent reason for signing a document that would commit her to an investment. She then acquires evidence that she is very tired and emotionally upset due to events that took place during the day – and that as a result, she cannot properly appreciate the force of her reasons. Hence, Ann has evidence that by signing the document she would be doing something she ought not to do, something she lacks reason to do. Raz seems to think that such evidence can make it the case that Ann ought not to sign. Or, consider discussions of moral uncertainty. Assume for the sake of argument that a Kantian theory is correct. I start out confident in the theory, and confident that in my current situation I ought to use the money on my savings account to help out a friend who has fallen ill. I then discover literature on utilitarianism, and become confident that the Kantian theory is wrong: it is morally wrong for me to waste my resources helping my friend, when the same money could save the lives of numerous people who lack access to proper nutrition and basic medicine. Is it still morally right for me to use my resources to help my friend; is that still what I overall ought to do? If the action of helping my friend would no longer have the status of being morally right as a result of this evidence – or as a result of my rational response to it – then it looks like we have a case of defeat of moral rightness by higher-order evidence that an action is not morally right.<sup>17</sup>

---

Lasonen-Aarnio 2015 for arguments against them). Even those who defend some form of *Rational Reflection* for credences could adopt *Higher-Order Defeat* (and a principle entailed by it, which I dub *Level-connection* below) as a constraint on rational *belief*. It is worth noting though that versions of *Rational Reflection*, together with a non-maximal credal threshold view of belief, clash with so-called *enkratic* principles very widely defended, both by proponents of higher-order defeat, and philosophers who reject versions of acquisition (see Lasonen-Aarnio 2015 for examples of such a clash, and Lasonen-Aarnio *forthcoming* for a critical discussion of enkratic principles). See also Dorst's essay in this volume.

<sup>16</sup> The discussion is in Chapter 1, "On Reasons for Action". My case of Ann is based on an example discussed by Raz. It is striking how much recent discussions in epistemology about how HOE forces one to 'bracket' some of one's first-order reasons resemble Raz's idea of excluding certain first-order reasons. Thanks to Arto Laitinen for drawing my attention to Raz's discussion.

<sup>17</sup> Numerous authors have defended norms that recommend different actions depending on what credences one assigns to various first-order normative theories. Such a norm might initially recommend acting as the Kantian theory recommends, but then switching as a result of becoming confident in utilitarianism. However, there is no agreement in the literature regarding just what status it is that these 'metanormative' theories are supposed to govern. Some views may be committed to the view that what it is *rational* to do depends on one's credences in claims about what it is *morally* right to do (e.g. Lockhart 2000, Ross 2006, Sepielli 2009). Gustafsson and Torpman (2014) couch their view as concerning 'what the morally conscientious agent chooses'. Guerrero (2007) speaks of blameworthiness or moral culpability. However, several of the views that have been discussed *are* committed to something very much like higher-order defeat of moral rightness, or defeat of the status of being what one overall ought to do. For instance, Gracely (1996) argues that one *should* act in accordance with the ethical theory most likely to be right. Harman (2011) argues against the view that 'an agent's moral uncertainty (and specific moral credences) are crucially relevant to how the agent should act', a view she attributes to various authors who put forth metanormative theories. Weatherson (2014) argues against a view on which what is morally right (or permitted, or wrong) is affected by moral



Hence, putative cases of defeat by higher-order evidence are not restricted to epistemology.

I will now describe an epistemic access condition I call *evincibility* – in particular, the evincibility of normative statuses such as rationality. I will then argue that *Higher-Order Defeat* entails commitment to the evincibility rationality.

### 3. Normative Evincibility

*Evincibility* is a kind of epistemic access. The access in question is a matter of certain truths always being sufficiently likely – likely at least to some positive degree  $r$  – on one's evidence. On the strongest construal,  $r = 1$ , and evincibility requires that these truths are always certain. Evincibility in the strongest sense entails that there can never be evidence bearing on the relevant class of truths. If we let  $r$  be some value very close to 0, then evincibility becomes very weak. Certain cases involving infinities aside, many find appealing a view on which evidence is factive, and the rational priors are never certain of falsehoods. While not equivalent to the weakest version of evincibility, such a view already guarantees that, setting infinite cases aside, all true propositions are probable to some degree above 0.

A rather natural, substantive form of evincibility assumes that  $r > .5$ : certain truths are always likely on one's evidence. When it comes to these truths, our evidence always points us in the right direction. Another, still rather substantive, thesis assumes that  $r \geq .5$ : certain truths can never be unlikely – our evidence can never point us in the wrong direction. But even if, for instance,  $r > .2$ , we still have what I am calling an evincibility thesis: the relevant truths cannot be very unlikely on the evidence.

In what follows, my concern will be with the evincibility of truths regarding the normative status of one's doxastic states (or other mental states, or actions); in particular, the evincibility of rationality. Questions about the evincibility of doxastic states themselves, and about their rational status, interact in various ways. For instance, if the rationality of one's states is evincible, then if one rationally believes  $p$ , it is fairly likely on one's evidence *that one believes  $p$*  (since it is fairly likely that one rationally believes  $p$ ). And of course, one kind of uncertainty about whether or not one rationally believes  $p$  derives from uncertainty about whether one believes  $p$  in the first place. In what follows, I will bracket uncertainty about what doxastic states one is in by assuming that one's doxastic states are strongly evincible: it is always certain on one's evidence whether or not one is in a doxastic state  $d$ , such as the state of believing  $p$ . While I think this assumption is false, it is helpful to set aside one kind of uncertainty in order to bring another kind into focus.

It will be helpful to talk of conditions: a condition can obtain at one time, and not obtain at another. For instance, *that I rationally believe that it is Monday* might be true on Monday, but no longer true on Tuesday. We can distinguish between a positive (what I meant by 'evincibility' above) and negative evincibility thesis. For instance, if we assume that the relevant threshold value  $r > .5$ , then according to the positive evincibility thesis, whenever a condition  $C$  obtains, it is likely that it obtains. According to the negative evincibility thesis, whenever  $C$  doesn't obtain, it is likely that it doesn't obtain. Together these entail that it is evincible *whether*  $C$  obtains: one's evidence always points to the truth regarding whether  $C$ . Note that evincibility can be characterized in a way that abstracts away from evidentialist assumptions: we could characterize it directly in doxastic terms instead. The idea would be that if a condition  $C$  is evincible, then one is in a position to be rationally confident at least to a relevant degree  $r$  that  $C$  obtains. Indeed, in what follows I will assume evincibility to have such doxastic ramifications.

---

uncertainty: if, for instance, the Kantian view is indeed correct, then the fact that I become convinced that effective altruism is the way to go does not make it morally wrong for me to help my friend.

*Normative evincibility* theses claim that truths concerning a relevant normative status of one's doxastic states (actions, etc.) are evincible. If rationality is (positively) evincible, then if one rationally believes  $p$ , one is in a position to be rationally confident at least to degree  $r$  that one's belief is rational – the condition *that I rationally believe  $p$*  is evincible. If rationality is both positively and negatively evincible, then one is always in a position to be reasonably confident about truths regarding the rational status of one's own beliefs.

According to a strong form of access internalism that most internalists these days shy away from, we always have a special kind of access to the normative (e.g. rational or justificatory status) of our doxastic states.<sup>18</sup> Evincibility is a way of spelling out what such access might involve: for instance, it may be a matter of the relevant propositions always being likely on one's evidence.

According to Williamson's (2000) use, a *luminous* condition is such that a subject is always in a position to know that it obtains when it does. Evincibility and luminosity feature distinct kinds of epistemic access. A proposition might be true and likely on one's evidence, even if one is in no position to know it because one is in a Gettier case. I also doubt whether being in a position to know a proposition  $p$  entails that  $p$  is likely on one's evidence: there are cases in which a proposition is unlikely on one's present total evidence, but one is nevertheless in a position to know it. (Perhaps this happens routinely when we form perceptual beliefs.) In any case, it should be clear that evincibility and luminosity make reference to different epistemic access conditions: that a proposition is likely at least to some degree  $r$  on the evidence is not tantamount to one being in a position to know it.<sup>19</sup>

At first sight, the evincibility of rationality might look to be at odds with the very considerations that motivate epistemologists to take higher-order evidence seriously. After all, examples of defeat involve even radically misleading evidence about the normative status of one's beliefs. Indeed, I argue below that there is a serious tension between *Acquisition* and evincibility. Nevertheless, rather than contradicting evincibility, *Higher-Order Defeat* in fact entails some form of the evincibility of rationality.

As many in the literature have realized, commitment to higher-order defeat is a commitment to some form of *level-connecting principle*. Indeed, many take commitment to such principles as a starting point, and argue for defeat on the grounds that they must be respected.<sup>20</sup> Christensen (2013: 90) remarks that level-connecting principles formulate "ways of taking beliefs in general to be rationally constrained by beliefs about what beliefs are rational". Sliwa and Horowitz (2015) lay down the desideratum that "One's rational first-order and higher-order doxastic attitudes should not be in tension", an example of such a tension being believing a proposition, while being merely 60% confident that the proposition is supported by one's evidence (and hence, being 40% confident that it is not). For instance, if *Higher-Order Defeat* states that evidence making it likely to some degree  $r$  or above that a state is irrational defeats its rationality, then HOD entails the following level-connection principle:

#### *Level-connection*

There is no case in which a relevant doxastic state  $d$  is rational but  $\Pr(d \text{ is irrational}) \geq r$ <sup>21</sup>

<sup>18</sup> See Chisholm (1988), Bonjour (1985), and for a more recent defence, Smithies (2012).

<sup>19</sup> Berker (2008) talks about *lustrous* conditions: a condition is lustrous iff whenever it obtains, one is in a *position to justifiably believe* that it obtains. The evincibility of a condition (given a way of specifying the threshold  $r$ ) may amount to its lustrousness, assuming that one is in a position to justifiably believe a proposition  $p$  just in case it is likely to at least degree  $r$  on one's evidence.

<sup>20</sup> E.g. Horowitz (2014), and Christensen in numerous writings, such as Christensen (2013).

<sup>21</sup> Does this principle entail that it is irrational to believe  $p$ , while suspending judgment about whether one's belief in  $p$  is irrational and hence, that a 'moderate' form of akrasia is always irrational? This is far from clear. Suspension of

If *Level-connection* was false, a state could be rational even if it was likely to degree  $r$  or above on one's evidence that it was not; hence, evidence making it likely to degree  $r$  or above that the state is irrational would not have defeating force with respect to the state. Note that though views endorsing higher-order defeat are committed to some kind of level-connecting principle, the converse is certainly not true. Indeed, various authors have defended level-connection principles, while denying at least some form of *Acquisition*.<sup>22</sup>

*Level-connection* in turn entails an evincibility claim. It states that it is impossible for evidence to make it likely to degree  $r$  or above that a state  $d$  is irrational, if  $d$  is in fact rational. But then, if  $d$  is rational, and the evidence always makes claims about its rationality or irrationality likely to some degree, it must be likely to some degree below  $r$  that  $d$  is irrational. In so far as a state is either rational or irrational, it follows that it must be likely to degree  $(1 - r)$  or above that  $d$  is rational. This is to claim that rationality is evincible; just how strong the claim is depends on exactly how *Higher-Order Defeat* is fleshed out.<sup>23</sup> I argued above that a natural value for  $r$ , one that would explain rather standard verdicts in cases of peer disagreement, is .5. Such a threshold would entail that rationality is evincible in the following sense: if a relevant doxastic state, such as the state of believing  $p$ , is rational, then it is *likely* (likely to some degree above .5) that it is rational. Alternatively, one might think that only evidence making it likelier than not that a state  $d$  is irrational defeats the rationality of the state. In that case, rationality would be evincible in the following sense: if a state is rational, it is not likely to degree .5 or above that it is irrational. That is, if a state is rational, it *cannot be unlikely* that it is rational.

In what follows, I will assume this latter view as a default. This thesis builds in a substantial assumption about defeat. But the assumption is in fact rather conservative. Moreover, my discussion below does not essentially rely on this threshold. Thus construed, *Higher-Order Defeat* entails the following thesis:

*Evincibility of rationality*

For all relevant doxastic states  $d$ , if one is in  $d$  and  $d$  is rational, then  $\Pr(d \text{ is rational}) \geq .5$

What the relevant doxastic states here are depends on one's commitments about defeat. I will assume that *Evincibility of rationality* applies at least to the state of full belief.

To block the inference from *Higher-Order Defeat* to evincibility, one could try to appeal to a distinction between two kinds of cases in which it is likely to degree .5 or above that one's belief in some proposition  $p$  is irrational. In the first case, one simply lacks reasons or evidence for thinking that one's belief in  $p$  is rational. In the second, one has positive reason to think that one's belief is irrational. One might propose that defeat only occurs in the second kind of case and hence, that no general evincibility thesis can be inferred from a commitment to defeat. (Note that according to a possible view, when one has no evidence bearing on whether a state is rational or irrational, propositions about its rationality simply have no probability on the evidence. But *that* is not the view being proposed: the idea is that it might be likely to degree

---

judgment might be appropriate in situations in which one has no evidence whatsoever bearing on the question of whether one's belief is irrational, situations in which the relevant evidential probabilities are imprecise or even undefined. Whether such 'moderate' forms of akrasia are irrational has been a source of debate. Hazlett (2012), for instance, argues that moderate akrasia is not irrational; others, like Feldman (2005), Huemer (2011), Smithies (2012), and Bergmann (2005), think that it is.

<sup>22</sup> E.g. Titelbaum (2015) and Littlejohn (2018).

<sup>23</sup> Perhaps one could allow for propositions about such higher-order matters to have no probability on the evidence whatsoever; in that case, we could formulate a conditional principle: if one's belief in a proposition  $p$  is rational, then in so far as the proposition *my belief in  $p$  is rational* has some probability on one's evidence, its probability must be above .5.

.5 or above that a belief is irrational simply because there is no evidence pointing to its rationality.)

However, level-connecting principles are completely silent on such issues regarding the evidential basis of one's attitudes, for they are structural in nature. If defeat is motivated by appeal to the idea that what doxastic states are rational is constrained by one's opinions about what doxastic states are rational – as it often is – then there is no room for the kind of distinction being drawn. From this perspective, believing  $p$  while being confident that it is irrational for one to believe  $p$ , for instance, is *always* bad. Moreover, the strategy discussed assumes that there can be cases in which it is .5 or above likely on a subject's evidence that a belief is irrational even though the subject lacks any evidence regarding the matter. On such a view there isn't a presumption in favour of our own rationality built into the rational priors – in fact, there might even be a presumption in favour of irrationality. But this raises tricky issues about how we could ever acquire evidence about our own rationality. According to a view able to avoid both scepticism and a kind of bootstrapping many would regard as dubious, we are entitled to assume our own faculties to function rationally, in the absence of evidence to the contrary. But if such views are right, then any case in which it is rational to be confident to some degree above .5 that a belief one holds is irrational is one in which there is some positive evidence for its irrationality. And then, there is no work left to do for the distinction appealed to.

I have argued that higher-order defeat commits one to some form of the evincibility of rationality. At this point the following worry might arise: aren't views committed to defeat by higher-order evidence committed to both affirming and denying the evincibility of rationality? On the one hand, one can acquire deeply misleading higher-order evidence bearing on the rational status of one's doxastic states (*Acquisition*). On the other, *Higher-Order Defeat* entails that the rational status of one's doxastic states is evincible: if a belief is rational, then it cannot be *too* unlikely that it is rational (in fact, according to the thesis I dubbed *Evincibility of rationality*, if a state is rational, it cannot be unlikely to be rational).

There is no blatant logical contradiction here – though, as I will argue, a deep tension remains. By *Acquisition*, one can *acquire* misleading evidence regarding the rational status of one's beliefs, but assuming *Defeat*, the result of acquiring such evidence is not a situation in which one is (radically) misled about the rational status of one's current beliefs. Consider again Rezi. At a time  $t$ , Rezi formed the belief that the appropriate treatment is a 10mg dose of Wellstrol ( $p_w$ ). At a subsequent time  $t'$ , she received the testimony of the neuroscientists. Given this new evidence, it is likely on her total evidence at  $t'$  that her belief in  $p_w$  was (and still is, assuming she holds it) irrational. However, if the rationality of the belief is defeated, then Rezi can no longer rationally believe  $p_w$ . Instead, perhaps it is appropriate for Rezi to suspend judgment about the matter. Hence, while Rezi does not have access to the rational status of her belief at the earlier time  $t$ , the rational status of her current doxastic state (say her suspension of judgment) might be evincible.<sup>24</sup>

In the rest of this paper, I will argue that though there is no straightforward logical contradiction, there is nevertheless a serious tension between *Acquisition* and *Evincibility of rationality*. There is considerable pressure on those who are liberal about the possibility of misleading evidence regarding the rational status of one's beliefs to give up on defeat by higher-order evidence.

---

<sup>24</sup> See also Skipper (*forthcoming*, section 4), who distinguishes between higher-order defeat, which is a diachronic phenomenon, and 'self-misleading evidence', which is synchronic. Having evidence that is self-misleading to a sufficiently high degree would be incompatible with evincibility, but as Skipper argues, defeat doesn't entail the possibility of self-misleading evidence.

#### 4. A tension

The dialectic of this section will be as follows. There are two distinct (though not exclusive) ways of being misled about the normative status of one's own beliefs and other doxastic states – and correspondingly, two different kinds of misleading HOE. Some HOE is misleading regarding what the correct theories or norms are; other HOE is misleading regarding contingent features of one's epistemic situation (and some HOE is misleading regarding both). As an example of the former, one might acquire evidence calling into question an evidentialist norm stating that one is permitted to believe  $p$  when one's evidence makes  $p$  likely; or, one might acquire evidence calling into question whether a determinate body of evidence  $e$  in fact makes  $p$  likely or not. As an example of the latter, one might acquire evidence calling into question exactly what one's evidence consists in. Either kind of misleading evidence can threaten the evincibility of rationality.

Focusing on these two distinct kinds of HOE creates two different contexts in which to investigate and test views committed to both *Acquisition* and *Higher-Order Defeat*. I will challenge the joint tenability of the two theses in both contexts. The methodological assumption will be that if neither context provides a fertile ground for the combination of views under discussion – views committed to a phenomenon of defeat by higher-order evidence – then neither will messier contexts involving HOE that falls under both kinds.

Let me begin by distinguishing between these two different kinds of HOE.

For any possible case, what I will call a *complete theory* tells one what doxastic states (if any) one is permitted or required to be in in that case. We could think of such complete theories as functions.<sup>25</sup> Let a *total set of doxastic states* be a set of doxastic states, each doxastic state (such as the state of believing  $p$ ) in the set being a doxastic attitude toward a single proposition, such that one is in that total set of states just in case one is in each state in the set, and in no other doxastic states. A complete theory takes possible epistemic cases as inputs, and gives sets of total sets of doxastic states as outputs. **T** *permits* believing  $p$  in a case  $c$  just in case **T** takes  $c$  to a set of total sets of doxastic states such that believing  $p$  belongs to *at least one* of these total sets of doxastic states. **T** *requires* believing  $p$  in  $c$  just in case **T** takes  $c$  to a set of total sets of doxastic states such that believing  $p$  belongs to *all* of these total sets of doxastic states. (Indeed, if a theory permits only one total set of doxastic state in  $c$ , it will take  $c$  to a set with only that total set of states as a member.) **T** *forbids* believing  $p$  in  $c$  just in case **T** takes  $c$  to a set of total sets of doxastic states such that believing  $p$  belongs to *none* of these total sets of doxastic states

Each complete theory is a theory of some normative notion, such as rational (rationally permitted or required) doxastic states, justified ones, or perhaps just of what doxastic states one is overall permitted (forbidden, required) to be in. Theories are sensitive to certain (epistemically) relevant features of cases; just what those features are depends on the nature of the theory. For instance, an evidentialist theory might only be sensitive to what one's total evidence in a case is. Other theories might be sensitive to other factors, like pragmatic stakes. We can derive norms from theories. For instance, if a theory requires one to believe  $p$  in all cases in which a condition  $C$  obtains, then the theory will entail the norm 'If  $C$ , believe  $p$ !' Conversely, we can infer facts about theories from the correct norms: if 'If  $C$ , believe  $p$ !' is a correct norm, then the correct theory will require believing  $p$  in all of the possible case in which  $C$  obtains.<sup>26</sup>

---

<sup>25</sup> Such complete theories are discussed by Titelbaum (2015: 263) and Skipper (*forthcoming*) in connection with their discussion of rational requirements.

<sup>26</sup> In what follows, I will assume that norms are informative *generalizations*. For instance, if a theory permits believing  $p$  in a particular case  $c$  (among many other cases), and condition  $C$  obtains *only* in that case, and in no others, then 'If  $C$ , believe  $p$ !' won't be general enough to count as a norm in the intended sense.

The **first** kind of misleading HOE is evidence that bears on what theories are correct in the first place. Assume that *If C, believe p!* is a correct epistemic norm, where C is a contingent condition that obtains in some cases and not in others. An example of this sort of evidence is evidence that this norm is incorrect: I might, for instance, acquire evidence that I am in fact forbidden to believe *p* when condition C obtains. Or, take the norm *If your evidence makes p likely, believe p!* Assume that I know all the contingent facts about my epistemic situation, including exactly what my evidence is. Nevertheless, I might acquire misleading evidence that *E*, my total body of evidence, does not make *p* likely. This is evidence that the correct theory does not permit believing *p* in situations in which *E* is my total evidence. By bearing on what complete theories are correct, the first kind of HOE bears on a domain of immutable and necessary normative truths.

By contrast, the **second** kind of misleading HOE bears on contingent features of one's situation, including facts about what doxastic states one is in, or how they came about. As a limiting case, an agent could be certain what the correct theory recommends in any possible case, but not know what the correct theory recommends for her, in virtue of being uncertain about epistemically relevant features of her situation. Given a norm of the form *If C, believe p!*, I will refer to C as its *application condition*. (Note that application conditions can concern the basing of one's beliefs – and norms that concern doxastic justification or rationality standardly do.) As a general rule, misleading evidence concerning whether the application conditions of (correct) norms obtain falls into this second class.<sup>27</sup> The higher-order evidence in *Resident* is like this: for a wide range of candidate norms governing doxastically justified or rational belief, Rezi gets evidence that her belief that a 10mg dose of Wellstrol is the appropriate treatment doesn't satisfy the conditions specified by those norms. Her evidence makes it likely, for instance, that her belief isn't appropriately based on sufficient reasons or evidence, that it isn't the output of a reliable cognitive process, etc.

I will now investigate the stability of views that endorse both *Acquisition* and *Higher-Order Defeat* in two different contexts, corresponding to these two different kinds of HOE that one might acquire. I will first bracket uncertainty about the kinds of necessary normative truths described above (e.g. truths concerning which norms are correct), focusing on uncertainty regarding contingent application conditions that arises from misleading HOE concerning one's epistemic situation; I will then bracket uncertainty about application conditions, and more broadly, contingent features of one's situation, focusing on uncertainty regarding what theories and norms are correct.

#### (i) *Uncertainty about application conditions of norms*

Focus first on misleading HOE that bears on contingent, epistemically relevant features of one's situation. It is clear that the evincibility of rationality puts constraints on the nature and strength of misleading evidence one could have regarding one's current situation. To see this, consider the following kind of case. I know (and it is likely on my evidence) that *Believe p if C!* is a correct norm; similarly for *Disbelieve p if C'!* Assume that I in fact believe *p*, and that C obtains. However, my evidence regarding my situation is misleading: it is likely on my evidence that C' obtains. Hence, it is likely that my belief is irrational – it is likely that I am in a situation in which norms

---

<sup>27</sup> This is just a general rule. Consider the norm 'Believe *p* just in case *p* is likely on your evidence!'. Evidence that is misleading regarding what evidence one has in the first place falls under the second kind of HOE concerning contingent matters. By contrast, evidence that is misleading regarding whether a determinate body of evidence *E* that one has makes *p* likely falls under the first kind of HOE: it is evidence regarding necessary truths concerning which theories are correct.

I know to be correct recommend disbelief. Then, *that my belief in  $p$  is rational* is not evincible: it is false, but likely on my evidence to be true.<sup>28</sup>

Demonstrating that *some* cases in which application conditions of norms fail to be evincible create failures of the evincibility of rationality isn't, of course, to show that the evincibility of rationality entails that the application conditions of norms must be evincible. Assume, for instance, that the correct norms deem it rational to believe  $p$  if *any* of the conditions  $C_1, \dots, C_n$  obtain. The evincibility of rationality is compatible with e.g.  $C_1$  obtaining, but having misleading evidence pointing to  $C_2$ . However, it is difficult to see a motivated way to allow having radically misleading evidence about central features of one's epistemic situation, while nevertheless retaining the evincibility of rationality. Consider, for instance, uncertainty about exactly what one's evidence is. Assume that my total body of evidence is  $E$ , and that  $E$  makes  $p$  likely. As a result, I am permitted (let us assume) to believe  $p$ . However, my total evidence is misleading regarding what my evidence is: it is likely that my total evidence is  $E^*$ . Assume that  $E^*$  still makes  $p$  likely. The rationality of my belief in  $p$  is evincible. So far, so good. But of course, this is not the only doxastic state I am in. If  $E^*$  is a different body of evidence from  $E$ , we should expect there to be *some* doxastic state  $d$  that is rational given one body of evidence, but not rational given the other. The rationality of  $d$  won't be evincible: it is rational, but likely not to be rational.

It is difficult to see how the evincibility of rationality could be retained in the current context unless the application conditions of correct epistemic norms were evincible. If our evidence could even radically mislead us regarding epistemically central features of our situations, counterexamples to the evincibility of rationality are bound to crop up. In this section, then, I will consider the following kind of view: one can acquire even radically misleading evidence regarding whether the application conditions of a correct epistemic norm obtain, but one cannot have (very) misleading evidence regarding such matters. I will first discuss the worry that application conditions of correct epistemic norms are not evincible, and a failed but instructive attempt to construct evincible conditions out of non-evincible ones. I then look at the strongest candidates for strongly evincible application conditions, conditions that concern a special class of internal mental states. I argue that views stipulating such states are hostile to higher-order defeat, and generalize the lesson: where appeal to a special class of internal mental states might help with the evincibility of rationality, it is at odds with instances of *Acquisition* relevant to putative cases of defeat by HOE.

A first objection to any view entailing that the application conditions of correct epistemic norms are evincible is that the application conditions of correct epistemic norms are *not* evincible. For pretty much any condition, the objection goes, our evidence about whether it obtains can be radically misleading. Consider, for instance, an evidentialist norm that requires believing  $p$  when  $p$  is sufficiently likely on one's evidence (possibly together with other conditions, such as that one has considered whether  $p$ , and/or that whether  $p$  is relevant in one's situation). Assume that evidence is factive: for a proposition  $p$  to be part of one's evidence,  $p$  must be true. Further, assume that for  $p$  to be part of one's evidence, one must bear some suitable epistemic relation to  $p$  (such as knowing  $p$ , or justifiably believing  $p$ ). Such commitments make it implausible to think that conditions such as *that  $p$  is likely on my evidence* are evincible in any very interesting sense. First, some proposition  $e$  might fail to be part of my evidence due to being false, even if it is likely that  $e$  is part of my evidence – and hence, it is likely that my evidence supports  $p$ , in virtue of the fact that  $e$  strongly supports  $p$ . Any theory on which evidence consists exclusively of true propositions gives rise to the possibility of such

---

<sup>28</sup> The argument assumed that I know the correct norms. There is an *ad hoc* way to block the argument: even if I could have misleading evidence regarding application conditions of norms, perhaps in such circumstances I would also always have misleading evidence regarding the correct norms, and the two mistakes would cancel each other out. I will set such views aside, as they appear to be egregiously *ad hoc*.

case.<sup>29</sup> Second, some proposition  $e'$  might be part of my evidence, even if it is likely that it is not: for instance, if there is such a thing as improbable knowing, I might in fact know  $p$ , even if it is very unlikely that I know  $e'$  and hence, unlikely that  $e$  is in fact part of my evidence. In such a situation my evidence might in fact make  $p$  likely, even if it is highly improbable that it makes  $p$  likely.

Such issues have gotten a decent bit of coverage in recent epistemology.<sup>30</sup> The point is that as far as the present project goes, acknowledging that the application conditions of epistemic norms are not evincible means that we have not found a way to jointly accommodate *Acquisition* and *Higher-Order Defeat*, for we haven't even found a way to accommodate *Higher-Order Defeat*.

One possible reply to such worries is that though *prima facie* plausible candidates for application conditions of true epistemic norms are not evincible, we can manufacture evincible conditions from non-evincible ones by adding a further condition acting as a kind of no defeater clause. For instance, assume that evidence making it at least .5 likely that a doxastic state is not rational is sufficient to defeat the rationality of that state (the precise threshold won't matter). The idea now is to construct application conditions that rule out having such defeating evidence.

Here is a first stab. On one kind of view, some sort of evidentialist condition is necessary and sufficient for the rationality of a belief: for a belief to be rational, it must be sufficiently likely on the evidence, and perhaps based in the right kind of way on evidence making it likely (let this be condition  $C$  – the reader can plug in any non-evincible condition in place of  $C$  below). What we want is an additional necessary condition that takes care of higher-order defeat: that  $C$  itself be likely to obtain on one's evidence. And hence, one might propose the following overall (necessary and sufficient) condition  $C^*$  on rationality:

- $$C^* = \text{I. } C$$
2. It is likely to some degree above .5 that 1.

Does this rule out the possibility of cases in which  $C^*$  obtains, but it is likely to degree .5 or above that  $C^*$  doesn't obtain? It doesn't.  $C^*$  is equivalent to the conjunction of 1. and 2. Consider a case in which  $C$  itself is likely, but condition 2. is not – and hence, though  $C$  is likely, it is not likely to be likely. This would be a case in which  $C^*$  obtains, even though it is likely to degree .5 or above that it does not and hence, that one's belief is not rational. But then, we would have a case in which *Higher-Order Defeat* is not respected.

At this point one might propose the following fix: simply require that 2. itself is likely, and likely to be likely, etc. That is:

- $$C^* = \text{I. } C$$
2. It is likely to some degree above .5 that 1. ( $\text{pr}(C) > .5$ )
  3. It is likely to some degree above .5 that 2. ( $\text{pr}(\text{pr}(C) > .5) > .5$ )
  - .
  - .
  - .

In a case in which each of these conditions (of which there is a countable infinity) obtains, we might say that  $C$  is *super-evincible*. But we are imagining that  $C^*$  constitutes the application

<sup>29</sup> Cf. Weatherson (2011).

<sup>30</sup> For instance, Christensen (2010b), Horowitz (2014), Lasonen-Aarnio (2015, *forthcoming*) Weatherson (2011), Williamson (2011, 2014), Worsnip (2018), Salow (2018).



condition of a correct epistemic norm. So it is  $C^*$  itself, and not merely  $C$ , that needs to be evincible.

But now we encounter essentially the same problem as before: even the obtaining of all of the above conditions 1., 2, ... does not entail that  $C^*$  itself is evincible – it does not entail that  $C^*$  is likely to some degree above .5.  $C^*$  is equivalent to the conjunction of its conditions; hence,  $C^*$  is likely just in case the conjunction of its conditions is likely. Each condition  $n \geq 2$  specifies that condition  $n - 1$  be likely. However, the fact that each conjunct individually is likely does not entail that the conjunction as a whole is likely. We have still not managed to specify an evincible condition.<sup>31</sup>

Lesson: It is not that easy to simply build up evincible conditions! Let me now explore a different tactic that does not give into anti-evincibility considerations in the first place.

Views on which the application conditions of correct epistemic norms are evincible to begin with offer a more promising platform for *Acquisition* and *Higher-Order Defeat*. Some argue that we always have perfect access to our evidence: if  $e$  is not part of my evidence, it is certain that  $e$  is not part of my evidence; and if  $e$  is part of my evidence, it is certain that  $e$  is part of my evidence.<sup>32</sup> But to meet the present challenge, one must show not only that application conditions of correct norms are evincible, but also that *Acquisition* holds: one can nevertheless acquire misleading evidence regarding these conditions. Those who think that we have perfect access to our evidence tend to think that evidence consists of a special class of internal mental states, or of propositions concerning such states. Let us see if appeal to such states would help.

Assume that the correct epistemic norms make reference to a domain of accessible internal mental states. Perhaps, for instance, they make reference to *seemings*, a *sui generis* class of mental states the obtaining of which is always certain on one's evidence – at least as long as one undergoes the relevant seeming. Assume that such seemings have a phenomenology with a special kind of epistemic *glow*: as long as one is actually in such a state, there can be no rational doubt that one is in it (compare such states with Descartes' clear and distinct ideas). Or: whenever one is in the state, it is evincible in a very strong way that one is in it. As a result, my present, occurrent seemings are perfectly evincible, but not my past seemings, for I can rationally doubt what seemings I experienced just a moment ago.<sup>33</sup> Even when there is no forgetting, this might explain how I could rationally doubt my past seemings, but not my present seemings. And in so far as the application conditions of correct epistemic norms only make reference to seeming-like states, this would seem to give the defeatist what she wants: I might

---

<sup>31</sup> There are frames of probabilistic epistemic logic we can use to model such cases (for more on such frames, see Williamson 2014). Consider a probabilistic frame  $\langle W, R, Pr \rangle$ , where  $W$  is a finite set (informally, of worlds),  $R$  is diadic relation over  $W$  (informally, a relation of accessibility between worlds, mapping a world  $x$  to all the worlds compatible with one's evidence at  $x$ ), and  $Pr$  is a prior probability distribution over subsets of  $W$ . Propositions can be modelled as subsets of  $W$ . Let  $W = \{w_1, w_2, w_3\}$ ;  $R(w_1) = \{w_1, w_2, w_3\}$ ,  $R(w_2) = \{w_1, w_2, w_3\}$ , and  $R(w_3) = \{w_2, w_3\}$ . Let  $Pr$  be uniform. A probability function  $Pr_x$  at world  $x$  results from conditionalizing  $Pr$  on the total evidence at that world – i.e. the set of accessible worlds. Note that in this frame  $Pr_{w_1} = Pr_{w_2} = Pr$ .

Let  $p$  be  $\{w_1, w_3\}$ . Let ' $pr(p) > .5$ ' stand for the 'the probability of  $p$  is above .5'. (It is important that I am here using ' $pr(p)$ ' for a definite description). So, for instance, in the frame assumed the proposition  $pr(p) > .5 = \{w_1, w_2\}$ . Let  $p$  (already defined above),  $p'$ , and  $p^*$  be as follows:

$$\begin{aligned} p &= \{w_1, w_3\}. \\ p' &= pr(p) > .5 \ \& \ pr(pr(p) > .5) > .5 \ \& \ pr(pr(pr(p) > .5) > .5) > .5 \dots = \{w_1, w_2\}. \\ p^* &= p \ \& \ p' = \{w_1\} \end{aligned}$$

At  $w_1$ ,  $p$  is "super-evincible" (it is likely, likely to be likely, etc), but  $Pr_{w_1}(p^*) = 1/3$ .

<sup>32</sup> For an interesting recent defence of such views, see Salow (2018).

<sup>33</sup> See Hawthorne and Lasonen-Aarnio (*forthcoming*) for a critical discussion of such views.

have misleading evidence about whether it ever seemed to me that  $p$  in the first place, and as a result, come to doubt whether my belief in  $p$  was ever rational to begin with.

However, for the purposes of explaining higher-order defeat, such epistemic glow is *too* powerful. The problem is that acquiring misleading HOE doesn't always march step in step with ceasing to be in the relevant seeming-states. Indeed, one of the peculiarities of defeat by HOE is that it calls for adjusting one's beliefs – or so many have argued – even if one's original evidence is still in place, and even if one correctly appreciates its force. So, for instance, it might still seem to me that  $p$ , even if I have evidence that it doesn't in fact seem to me that  $p$ . But as long as it still seems to me that  $p$ , and my seemings *glow*, we don't have a case in which it is rational for me to doubt whether it ever seemed to me that  $p$  in the first place. The problem is that the relevant application condition making reference to seemings might still obtain. Glow might explain evincibility, but doesn't give the defeatist what she wants, for it is impossible to acquire misleading evidence about one's seemings so long as one continues to experience them. (And of course, the strategy is hopeless if these special kinds of states that glow don't exist in the first place.)

The lesson can be generalized. Whether what is at issue is a very strong kind of evincibility on which certain conditions are always certain to obtain, or a weaker kind on which they cannot be likely *not* to obtain, the most plausible candidates for evincible conditions are conditions making reference to a special domain of internal mental states with a glow-like property. Glow guarantees that our overall evidence regarding these special states cannot be (too) misleading. Let  $C$  be a condition stating that one is in such an internal state, such as the condition *that it seems to one that  $p$* . The problem now is that in order for one to acquire evidence making it likely that  $C$  doesn't obtain,  $C$  has to *cease* to obtain. But if  $C$  concerns the obtaining of internal mental states, there is no reason to think that it always ceases to obtain when higher-order evidence that  $C$  doesn't obtain comes in – and as a result, there is no reason to think that the obtaining of  $C$  marches in step with common verdicts about higher-order defeat.

I have not proved that it is *impossible* to accommodate systematic higher-order defeat – and in particular, its commitment to both the evincibility of rationality and *Acquisition* – within the present context that allows for misleading evidence regarding contingent features of one's epistemic situation, but not regarding what theories or norms are correct. But I take the considerations given to at least give us strong *prima facie* reasons to be doubtful. What is needed is norms with application conditions that satisfy the following constraint: one can *acquire* even radically misleading evidence about whether the conditions obtain, but one cannot *have* evidence that is too misleading regarding whether they obtain. A first challenge for such views is to show that the application conditions of correct epistemic norms are evincible in the first place. But even setting this aside, a serious problem remains: where evincibility pushes one to think that the application conditions of correct epistemic norms make reference to a special class of internal states, *Acquisition* creates pressure in the opposite direction.

Perhaps not surprisingly, several proponents of higher-order defeat and/or of level-connecting principles have explicitly acknowledged that the application conditions of norms are not evincible – and that their failure to be evincible creates seeming counterexamples to level-connecting principles, such as ones prohibiting a kind of epistemic akrasia.<sup>34</sup> Let me now investigate whether normative uncertainty creates a more fertile ground for higher-order defeat.

---

<sup>34</sup> E.g. Christensen (2010b), Titelbaum (2015), Horowitz (2014), Worsnip (2018)

(ii) *Normative uncertainty and evincibility*

I will now bracket uncertainty about and misleading evidence bearing on the application conditions of norms, and more generally, on contingent features of one's epistemic situation – including facts about one's own mental states and processes – in order to investigate evidence bearing on the rationality of one's doxastic states in virtue of bearing on necessary normative truths.

*Acquisition* states that it is possible to acquire evidence making it rational to believe, or be confident, in falsehoods regarding the rational status of one's doxastic states. Here is what acquiring such evidence would involve in the current context. Assume that I am in a case *c*, and that a doxastic state *d* is permitted by the correct theory/norms in *c*. Acquiring new HOE could then land me in a case *c'* in which it is rational for me to believe (or be confident) in the falsehood *I am forbidden to believe p in c*. One way this could happen is if I acquire evidence making a false (complete) theory **T<sub>FALSE</sub>** likely, where **T<sub>FALSE</sub>** forbids believing *p* in *c*. Another, more realistic option, is acquiring evidence making likely a possibly large *disjunction* of false theories, all of which forbid believing *p* in *c*. This might happen by acquiring evidence for false norms. I might, for instance, acquire evidence that *d* is only permitted when some condition *C* obtains, a condition I know not to obtain in *c*.

The evincibility of rationality, however, is incompatible with having evidence making likely such false normative claims *about my current situation*: it cannot be likely on my evidence that being in some state *d* is forbidden in my current situation, if it is in fact permitted. To see this, consider the following kind of case. Assume that I am in fact required to disbelieve *p* (that's what the correct theory/norms urge me to do), and I do so. However, I have misleading evidence regarding what the correct norms are. It is very likely on my evidence that I am required to believe *p* if *C*, for some condition *C*. Assume that now *C* obtains, and that I know (and it is likely on my evidence) that *C* obtains. Then, the rationality of my state of disbelieving *p* is not evincible: it is by assumption rational, but it is very likely on my evidence that I am rationally required to believe *p* instead and hence, that disbelieving *p* is rationally forbidden. More generally, given that we are bracketing possible uncertainty about contingent features of one's epistemic situation, the evincibility of rationality will be threatened in any case *c* in which one is permissibly in some doxastic state *d*, but has evidence for a false theory, or for a false norm, that forbids being in *d* in *c*.

In the current context, higher-order defeat commits one to the view that while I can be even radically misled regarding a range of normative truths that are about situations other than the one I am currently in, I cannot be thus misled about my current situation. Here is a quick argument that such a view is a non-starter:

'If the rational status of my (current) beliefs is always evincible, then a range of normative truths must be sufficiently likely at all times and hence, evincible. But then, I cannot acquire misleading evidence regarding such truths, evidence that could trigger defeat by higher-order evidence.'

If the evincibility of rationality required the evincibility of the complete normative theory **T<sub>TRUE</sub>**, or the evincibility of all of the correct norms, then this reasoning would be correct. For then, that complete theory, or the total set of true norms, would have to be sufficiently likely in any possible case, and hence, evincible. However, it is not clear whether the evincibility of rationality does require the evincibility of complete normative theories, or of all correct norms. To see this, assume that I am now in case *c*. Because rationality is evincible, at least theories or norms forbidding some of my perfectly rational doxastic states cannot be too likely on my evidence. A range of truths stating facts about what doxastic states the correct theories or norms require or permit in *c* must be sufficiently likely. It doesn't follow from this, however, that the unique complete true theory **T<sub>TRUE</sub>**, or the total set of correct norms, must be

sufficiently likely on my evidence. After all, lot of the recommendations made by the complete theory apply to cases other than the one I am in. As I transition to another case  $c'$ , I could acquire even radically misleading evidence regarding normative facts pertaining to  $c$ , as long as I don't have the kind of misleading evidence regarding what is now my actual case,  $c'$ , that would threaten the evincibility of rationality. *In principle* both *Acquisition* and the evincibility of rationality could be retained, as long as my evidence bears on normative matters in the right kinds of ways as I transition from one epistemic situation to another.

Nevertheless, such a view raises some serious explanatory challenges.

If I always have a kind of access to facts about what the correct normative theory recommends in my current situation, why don't I have access to facts about what it recommends in other situations? The purported asymmetry in one's access to a range of normative truths regarding the situation one currently occupies, *versus* other situations, needs to be explained. Mike Titelbaum (2015) has recently argued that the challenge cannot be met. We need some epistemic story of how it is that it can never be justified for one to both hold a doxastic state  $d$ , and to (falsely) believe that  $d$  is rationally forbidden. But, Titelbaum argues, any plausible story will generalize to story banning such normative mistakes regarding other situations. For instance, an obvious explanation for why one could never have evidence making it likely that a doxastic state  $d$  is forbidden, when  $d$  is in fact permitted, is that "every agent possesses *a priori*, propositional justification for true beliefs about the requirements of rationality in her current situation". But if one can gain true beliefs about what the correct theory requires or forbids in one's current situation just by reflecting on features of what situation, then surely one can gain such true beliefs regarding *other* situations. Conversely, one could argue that a view explaining how it is that we *don't* have access to what the correct theory recommends in situations other than our own will generalize to an explanation of why we don't have such access to our own situation.

Here is another way to bring out the peculiar nature of the view under discussion. Assume that you are traveling with a testifier who makes claims about the local customs and habits at various places and times. The testifier occasionally lies. However, whenever she is lying about the place you are both currently in, she always gives out some telltale sign that she is lying. You need never be misled about the customs and habits in your current location by such a testifier. Analogously, consider an epistemic theory that makes false recommendations in some situations. However, assume that in all of the situations in which the theory makes false recommendations, there are reasons to reject the theory anyway – the theory is unlikely on the overall evidence. Or, at the very least, in none of these situations is the theory likely. Call such a theory *telltale* false. There is no worry of being misled by such a theory into holding false beliefs about the rational status of one's doxastic states, as the theory can never be likely in situations in which it strays from the right recommendations.

In the current context, a commitment to higher-order defeat (*Acquisition* + *Evincibility of rationality*) incurs commitment to the following claim: all *possible* epistemic theories are false in a telltale way – at least when it comes to their falsely claiming of some permitted doxastic state that it is forbidden. That is, some possible epistemic theories falsely claim of some situations that doxastic states that are in fact permissibly held in those situations are forbidden, *but in those situations those theories are never likely to be true*. But why are only theories that are false in a telltale way even *possible* in the first place? Assume that it is likely on my evidence that certain doxastic states are forbidden in cases  $s_1, \dots, s_n$ , even though those states are in fact permitted in those situations (I have acquired misleading HOE regarding those situations). There must be some guarantee that *if* I then transition to one of these epistemic situations, those normative claims cannot be likely. That is, it must either be impossible for me to come to be in one of these situations, or the relevant normative claims must automatically become unlikely once I come to be in them. But why might this be?

Let me describe an example that brings out these challenges.

Let  $T_E$  be an evidentialist theory stating that one is rationally permitted to believe a proposition  $p$  whenever  $p$  is at least .9 likely on one's evidence.  $T_{E+S}$  is a theory that allows pragmatic stakes to affect just how likely a proposition must be on one's evidence in order for one to be permitted to believe it. Assume that according to  $T_{E+S}$ , when the stakes are high,  $p$  must be at least .99 likely in order for a belief in  $p$  to be rationally permitted; if  $p$  is only .9 likely, the rational attitude is suspension of judgment. (These are simplified toy theories, of course.) Assume for the sake of argument that  $T_{E+S}$  is false, and that  $T_E$  is true: stakes can have no such effect on what opinions are rational.

Here is my situation at 1.50pm: the stakes are low, and I rationally believe that there is a 2pm train that will take me to the hospital to see my aunt (proposition *train*), on the grounds that it is a weekday, and I took the 2pm train on a weekday a couple of weeks ago. *Train* is in fact at least .9 likely on my total evidence and hence, my belief is rational. However, I have misleading philosophical evidence, and at 1.50pm I am also rationally confident (perhaps I even rationally believe) that  $T_{E+S}$  is true – that is, that stakes can affect what it is rational for me to believe. Nevertheless, at 1.50pm I satisfy *Evincibility of rationality*, at least as far as my belief in *train* goes. As I am walking to the station, at 1.55pm I receive a phone call: I am told that my aunt is in critical condition. The stakes have now gone up, and I know it: according to  $T_{E+S}$ , it is no longer rationally permitted for me to believe *train*. Assume that I nevertheless continue to believe it. Given that it is still at least .9 likely and the fact that  $T_E$  is the true theory, my belief continues to be rational. However, the only way for me to continue to satisfy *Evincibility of rationality* is if I suddenly lose my confidence in  $T_{E+S}$ . But there is nothing that would make this doxastic shift rational: I don't acquire any new evidence bearing on such normative matters. Possible cases like this spell trouble for the evincibility of rationality.

In the case described, I start out confident in a theory that makes a false recommendation in situations other than my own. But I then come to be in one of these situations: what starts out as normative uncertainty regarding other epistemic situations or cases becomes uncertainty about my current situation.  $T_{E+S}$  is not false in a telltale way. But according to the present view, such theories are *impossible*. As far as I can see, there are two general, non *ad hoc* ways of block cases like the one I described from arising, and to meet the explanatory challenges raised above. The first is commitment to a picture of epistemic normativity entailing that if the stakes-sensitive theory  $T_{E+S}$  is likely on my evidence, its recommendations are *thereby* correct. Hence, the stipulation that the purely evidentialist theory  $T_E$  could still make a true recommendation in a situation in which  $T_{E+S}$  is likely to be true is simply false. This is a deeply troubling picture of epistemic normativity. The second option posits epistemic dilemmas in all of the problem situations in which evincibility is threatened. In the above example, though  $T_E$  could still be true, if a false theory  $T_{E+S}$  making a different recommendation is likely, there is *nothing* I could do that is rational.

Before I discuss these options, let me address the worry that I have taken uncertainty about necessary normative truths too seriously. Surely pretty much any view put forth for dealing with defeat by higher-order evidence has its restrictions, and is not tailored to accommodate just *any* evidence regarding normative matters. Consider the history of conciliatory views of peer disagreement. It was soon noticed that conciliatory views have untoward consequences if applied to disagreements concerning conciliationist views.<sup>35</sup> In response, Elga argues that the application of conciliationism should be restricted in some motivated way. But rather than establishing the joint tenability of *Acquisition* and *Higher-Order Defeat* (and hence, evincibility), Elga's restriction makes it impossible to acquire misleading evidence about normative matters. The kinds of complete theories I have discussed are what

---

<sup>35</sup> See Weatherson (2007), Elga (2010).

and Lewis (1971) calls *inductive methods*: systematic ways ‘of letting the available evidence govern your degree of belief in hypotheses’ – or even more generally, systematic ways of letting relevant features of one’s epistemic situation govern one’s doxastic states. Following Field (2010), Elga argues that all acceptable inductive methods must be coherent, which requires that they never recommend other methods over themselves. Any candidate for a true theory **T** must always recommend being confident in **T**. Indeed, Field (2010: 133) concludes on this basis that at least we must treat our basic inductive rules or methods as empirically indefeasible and hence, fully *a priori*. That is, we cannot rationally regard anything as evidence against them. While such a view is amenable to *Evincibility of rationality*, it is not amenable to the possibility of acquiring misleading higher-order evidence regarding what rationality consists in.

Let me now discuss two responses to the challenges raised.

### *Reply 1: Perspectivalism*

If I always have a kind of access to a range of facts about what the correct normative theory recommends in my current situation, why don’t I have access to facts about what it recommends in other situations? Relatedly, if a false theory **T<sub>FALSE</sub>** forbidding some doxastic states that are in fact permitted in some situations could be likely on my evidence, what is to prevent it from being likely in one of the situations in which it makes a false recommendation – in particular, in which it forbids being in doxastic states that are in fact permitted? That is, why are only theories that are false in a telltale way even *possible*?

There is a story answering these question, albeit one that paints a deeply troubling picture of epistemic normativity.

Consider a view on which the normative truths themselves are malleable, being shaped by one’s evidence or rational opinions regarding what the true theory or norms are. The idea is that if a theory is likely on my evidence, then its recommendations are *thereby* correct.<sup>36</sup> Such a view explains why all possible false theories are false in a telltale way: as a theory becomes unlikely, its recommendations *thereby* cease to be correct. The view would also answer the explanatory challenge raised by Titelbaum: there is a constitutive kind of dependence between which normative claims are correct, and which claims are likely on my evidence at a given time. The fact that such access is constitutive of their truth is what explains why I always have a kind of access to a range of normative truths that bear on what it is rational for me to believe in my current situation.

Of course, given that I could in principle have evidence for just about any theory – at least any of the theories ever put forth by epistemologists – this leads to a view on which just about anything can be made to go, as long as one has evidence for a theory on which it is correct. Littlejohn (2018) rightly calls the view a kind of epistemic anarchy. I won’t spend more time arguing against the view; I only note that it is an option, albeit one that very few would want to endorse.

The other way of meeting the challenges presented above posits epistemic dilemmas in any case in which the evincibility of rationality (and the relevant level-connecting principle) is threatened.

### *Reply 2: Dilemmas and evincibility*

Here is a procedure for turning any initial theory with the kinds of problems pointed to into one that both respects *Evincibility of rationality*, and accommodates normative uncertainty.

---

<sup>36</sup> Littlejohn (2018) calls such a view *perspectivalist*, since on it facts about rationality depend on one’s perspective on rationality.

Consider all possible cases, and the recommendations made by a candidate true theory such as  $T_E$  in those cases regarding a particular proposition  $p$ . We can split these into 2 classes.

1. First, there are cases in which  $T_E$  permits or requires some doxastic attitude to  $p$ , and it is sufficiently likely on one's evidence that this recommendation is correct ( $T_E$  itself, some other theory making the same recommendation, or a disjunction of theories all making this recommendation, is sufficiently likely).
2. Second, there may be cases in which  $T_E$  permits some doxastic attitude to  $p$ , but it is *not* sufficiently likely on one's evidence that this recommendation is correct.

The first kinds of cases are unproblematic from the perspective of *Evincibility of rationality*. Here is a strategy for dealing with cases of the second kind, which threaten evincibility: simply rule that in them *no* state is rational, for they constitute kinds of epistemic dilemmas.

Why dilemmas? Because rationality is governed by two ideals that sometimes conflict. The first kinds of ideals are incorporated into our initial theory  $T_E$ , which I have for the sake of argument assumed to be an evidentialist theory. The second kinds of ideals have to do with taking into account higher-order evidence; one might refer to them, as Christensen does, as ideals of epistemic modesty. It is this second kind of ideal which prohibits believing  $p$  when believing  $p$  is not likely to be rational. Situations in which the two ideals come into conflict are epistemic dilemmas: nothing the subject could do in them is rational.<sup>37</sup>

Unlike the view just discussed on which anything goes (as long as it is likely to be right), the present view has no such implications. It does, however, have a similar negative implication: anything could be made *not* to go by evidence for a false theory. Are the consequences of this view plausible? Is it plausible, for instance, that just because a philosopher is unlucky enough to be inundated by evidence for an epistemological theory making a false recommendation in her current situation, she is doomed to irrationality, no matter what she does? Or, consider the analogous view in the moral case. Assume that the correct moral theory tells me to financially help my friend so she can be treated for cancer. I am, however, under the grips of consequentialist views that require me to help strangers in ways that are more effective, saving numerous lives with the same money. Assuming this recommendation is false, do we really want to say that by helping my friend I fail to act as I ought, since there is *nothing* I could do that would constitute acting in the morally right way?

Of course, friends of higher-order defeat don't *merely* make the negative claim that evidence that a belief is irrational defeats its rationality; they typically make positive recommendations about how one ought to take such evidence into account. For instance, Rezi should downgrade her confidence, or suspend judgment, in light of her HOE. This is even true according to Christensen, who advocates a version of the dilemma view. But a structure with conflicting ideals or norms cannot yield such overall recommendations: something more is needed, either a meta-rule, or some way of weighting the conflicting rule.

My purpose here is not to argue against the dilemma view (see Lasonen-Aarnio 2014), but to outline the options. Note that appeal to such dilemmas is not a way of giving up commitment to evincibility: it is still true that any situation in which it is rational to believe  $p$  is one in which it is sufficiently likely on one's evidence that it is rational to believe  $p$ . For any theory, the recommendations made by that theory can be correct only if it is sufficiently likely on one's evidence that they are correct. So reasons to think that rationality is not evincible are reasons to think the dilemma view cannot be correct.

---

<sup>37</sup> Christensen (2007a, 2010a) defends such dilemmas.

## Conclusions

I applaud taking seriously the idea that there can be deep uncertainty about the normative status of our own beliefs, intentions, and actions. We are fallible, even regarding how rational our beliefs are, and how morally right our actions are. Not all cases of normative success are ones in which we have access to the fact that we succeeded, and not all cases of normative failure are ones in which we have access to the fact that we fail. But numerous authors, in both epistemology and ethics, have moved from this realization to an endorsement of a systematic kind of higher-order defeat: if I am sufficiently rationally confident that I fail by some normative standard, then I *do* in fact fail by that standard. Norms urging us to downgrade our opinions in light of evidence about our failure, or to split the difference with our epistemic peers, are often presented as requirements of a proper kind of epistemic modesty.

I have argued that a commitment to systematic defeat by higher-order evidence amounts to a commitment to a kind of epistemic access, to what I have called normative evincibility. There is a sense in which such evincibility is not at all modest, as it denies that we could be deeply misled about how well we are doing when it comes to normative standards like rationality. Some authors like Smithies (2012) and Titelbaum (2015) happily endorse such access internalism at the outset. However, such views don't sit well with those who admit the possibility of deeply misleading higher-order evidence about normative matters. I have argued that there is at least a deep *prima facie* tension between two theses that those who endorse a systematic phenomenon of defeat by higher-order evidence are committed to, *Acquisition* and *Higher-Order Defeat*. If this is right, then we are left with two stable views: we can either endorse *Evincibility of rationality*, or endorse *Acquisition*, and give up on the hope of accommodating a systematic kind of defeat, or of formulating a level-connection principle not susceptible to counterexamples.

To be sure, none of the arguments given above are conclusive. But at the very least, I hope to have identified a key challenge for those committed to defeat by higher-order evidence: explaining the kind of asymmetry such views are committed to when it comes to one's epistemic access to a range of normative facts regarding one's own epistemic situation, versus a lack of such access to a range of normative facts regarding situations other than one's own.<sup>38</sup>

## References

Berker, Selim

(2008) "Luminosity Regained", *Philosophers' Imprint* 8/2: 1–22

Bergmann, Michael

(2005) "Defeaters and Higher-Level Requirements", *The Philosophical Quarterly*, Vol 55. 220: 419–436.

(2006) *Justification without Awareness*, Oxford University Press.

Bonjour, Laurence

(1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press

---

<sup>38</sup> I am immensely grateful for written comments on an earlier draft of this paper from the editors, a referee, Anna-Maria A. Eder, Mike Titelbaum, and Teru Thomas. Many thanks to discussions with Ville Aarnio, Jaakko Hirvelä, Mattias Skipper, and the audience at the *Bochum Epistemology Seminar* in June 2018



Chisholm, Roderick M.

(1977) *Theory of knowledge*, 2nd edition, Englewood Cliffs, NJ: Prentice-Hall.

Christensen, David

(2013) "Epistemic Modesty Defended", in Christensen and Lackey (eds.), *The Epistemology of Disagreement: New Essays*, Oxford University Press, pp. 77-97.

(2010a) "Higher-Order Evidence," *Philosophy and Phenomenological Research*, 81.1: 185-215.

(2010b) "Rational Reflection", *Philosophical Perspectives*, 24: 1, pp. 121-140.

(2007a) "Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals," *Oxford Studies in Epistemology* 2, pp. 3 - 31.

(2007b) "Epistemology of Disagreement: the Good News," *Philosophical Review* 116, pp. 187 – 217.

Dorst, Kevin

(forthcoming) "Evidence: A Guide for the Uncertain", *Philosophy and Phenomenological Research*.

Elga, Adam

(unpublished) "Lucky to be rational"

<http://www.princeton.edu/~adame/papers/bellingham-lucky.pdf>.

(2010) "How to Disagree About How to Disagree"

(2013) "The puzzle of the unmarked clock and the new rational reflection principle", *Philosophical Studies* 164.1: 127-139.

Feldman, Richard

(2005) "Respecting the evidence", *Philosophical Perspectives*, Volume 19, Issue 1, pp.95–119.

Field, Hartry

(2000) "A priorism as an evaluative notion", in *New Essays on the A Priori*, Boghossian and Peacocke (eds.), Clarendon Press, Oxford, pp. 117-148.

Gustafson, Johan and Torpman, Tom

(2014) "In Defence of My Favourite Theory", *Pacific Philosophical Quarterly*, 95.2: 159-174.

Gracely, Edward J.

(1996) "On the Noncomparability of Judgments Made by Different Ethical Theories", *Metaphilosophy* 27: 327–32.

Harman, Elizabeth

(2011) "Does Moral Ignorance Exculpate?", *Ratio*, XXIV.4: 443-468.

Hawthorne, John and Srinivasan, Amia

(2013) "Disagreement Without Transparency – Some Bleak Thoughts", in Christensen and Lackey (eds.), *The Epistemology of Disagreement: New Essays*, Oxford University Press, pp. 9-30.

Hazlett, Allan

(2012) "Higher-Order Epistemic Attitudes and Intellectual Humility", *Episteme* 9.3: 205-223.

Horowitz, Sophie

(2014) "Epistemic Akrasia", *Noûs* 48.4: 718-744.

Huemer, Michael

(2011) "The Puzzle of Metacoherence", *Philosophy and Phenomenological Research*, 82(1): 1-21.

Kelly, Tom

(2010) "Peer Disagreement and Higher-Order Evidence", in Feldman and Warfield.

Lackey, Jennifer

(2010) "What Should We Do when We Disagree", in T. S. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology III* (Oxford: Oxford University Press), 274-293.

Lasonen-Aarnio, Maria

(2010) "Unreasonable knowledge", *Philosophical Perspectives*, Vol 24:1, pp. 1-21.

(2013) "Disagreement and evidential attenuation", *Noûs* 47.4: 767-794.

(2014) "Higher-order evidence and the limits of defeat", *Philosophy and Phenomenological Research*, 88.2: 314-345.

(2015) "New Rational Reflection and Internalism about Rationality", *Oxford Studies in Epistemology* 5, pp. 145-179. 2015.

(forthcoming) "Enkrasia or Evidentialism?", *Philosophical Studies*.

Lewis, David

(1971) "Immodest Inductive Methods", *Philosophy of Science* 38(1): 54-63.

Lockhart, Ted

(2000) *Moral Uncertainty and Its Consequences*, Oxford: Oxford University Press.

Littlejohn, Clayton

(2018) "Stop Making Sense", *Philosophy and Phenomenological Research*, 96.2: 255-513

MacAskill, William

2014 *Normative Uncertainty*, DPhil Thesis, University of Oxford

Raz, Joseph

(1975) *Practical Reason and Norms*, Oxford University Press.

Ross, Jacob

(2006) "Rejecting Ethical Deflationism", *Ethics* 116: 742-768.

Salow, Bernhard

(2018) "The Externalist's Guide to Fishing Compliments", *Mind* 127.507: 691-728

Schechter, Joshua

(2011) "Rational Self-Doubt and the Failure of Closure", *Philosophical Studies* 163.2: 429-452.

Sepielli, Andrew

(2009) "What to Do When You Don't Know What to Do", *Oxford Studies in Metaethics* 4:5-28

Skipper, Mattias

(Forthcoming), "Reconciling Enkrasia and Higher-Order Defeat", *Erkenntnis*.

Sliwa, and Horowitz, Sophie

(2015) "Respecting all the evidence", *Philosophical Studies*, 172.11: 2835-2858.

Smithies, Declan

2012 "Moore's Paradox and the Accessibility of Justification", *Philosophy and Phenomenological Research* LXXXV: 2, 273-300.

Titelbaum, Michael G.

(2015) "Rationality's Fixed Point (Or: In Defence of Right Reason)", *Oxford Studies in Epistemology* 5: 253-294.

Worsnip, Alex

(2018) "The Conflict of Evidence and Coherence", *Philosophy and Phenomenological Research* 96.1: 3-44.

Weatherson, Brian

(2007) "Disagreements, Philosophical and Otherwise", in Christensen and Lackey (eds.), *The Epistemology of Disagreement: New Essays*, Oxford University Press, pp. 54-75.

(2011) "Stalnaker on Sleeping Beauty", *Philosophical Studies*, 155, pp. 445-56.

(2014) "Running Risks Morally", *Philosophical Studies* 167 (1): 141-63.

Van Wietmarschen, Han

2013 "Peer Disagreement, Evidence, and Well-Groundedness", *The Philosophical Review*, 122.3: 395-425.

Williamson, Timothy

(2000) *Knowledge and its Limits*, Oxford University Press.

(2011) "Improbable knowing", In T. Dougherty (ed.), *Evidentialism and its Discontents*, Oxford University Press.

(2014) "Very improbable knowing", *Erkenntnis*, 79: 971-999.